UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/626,984	07/25/2003	Kenji Kawai	013.0207.US.UTL	4146

22895 7590 11/27/2007
CASCADIA INTELLECTUAL PROPERTY
500 UNION STREET
SUITE 1005
SEATTLE, WA 98101

EXAMINER	
HARPER, LEON JONATHAN	
ART UNIT	**PAPER NUMBER**
2166	

MAIL DATE	DELIVERY MODE
11/27/2007	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Office Action Summary	Application No.	Applicant(s)	
	10/626,984	KAWAI ET AL.	
	Examiner	Art Unit	
	Leon J. Harper	2166	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *07 September 2007*.

2a)☒ This action is **FINAL**. 2b)☐ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-6,9-14,17-27,31,35-44,48,52 and 53* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-6,9-14,17-27,31,35-44,48,52 and 53* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☒ Information Disclosure Statement(s) (PTO/SB/08) Paper No(s)/Mail Date *10/1/07*.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .
5) ☐ Notice of Informal Patent Application
6) ☐ Other: _____.

DETAILED ACTION

Response to Amendment

1. The amendment filed 9/7/2007 has been entered. Claims 7, 8, 15, 16, 28-30, 46, and 47 have been cancelled. Claims 1, 9, 17, 18, 31, 35, 48, 52, and 53 have been amended. Accordingly, claims 1-6, 9-14, 17-27, 31, 35-44, 48, 52, and 53 are pending in this office action.

Claim Rejections - 35 USC § 103

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

The factual inquiries set forth in *Graham* **v.** *John Deere Co.*, 383 U.S. 1, 148 USPQ 459 (1966), that are applied for establishing a background for determining obviousness under 35 U.S.C. 103(a) are summarized as follows:

1. Determining the scope and contents of the prior art.
2. Ascertaining the differences between the prior art and the claims at issue.
3. Resolving the level of ordinary skill in the pertinent art.
4. Considering objective evidence present in the application indicating obviousness or nonobviousness.

Claims 1-6, 9-14, 17-23 ,35-40, 52,53 are rejected under 35 U.S.C. 103(a) as being unpatentable over Wo 03060766 (hereinafter Lind) (art of record) in view of US 6560597 (hereinafter Dhil).

As for claim 1 Lind discloses: a scoring module determining a score which is assigned to at least one concept that has been extracted from a plurality of electronically-stored documents (See page 17 lines 20-24 note: definition of document corpus) wherein the score is based on at least one of a frequency of occurrence of the at least one concept within at least one such document, a concept weight, a structural weight, and a corpus weight, forming the score assigned to the at one concept as a normalized score vector for each such document, and determining a similarity between the normalized score vector for each such document as an inner product of each normalized score vector (See page 30 line 30- page 31 line 1).

; (See page 7 lines 20-24) a clustering module forming clusters of the documents comprising a selection sub module evaluating a set of candidate seed documents selected from the plurality of documents; a seed document identification submodule identifying a set of seed documents by applying the similarity as a best fit to each such candidate seed document; a non-seed document identification submodule identifying a plurality of non-seed documents; a comparison submodule determining the similarity between each non-seed document and a center of each cluster; and a clustering submodule grouping each such non-seed document into a cluster with the best fit, subject to a minimum fit See page 28 line 8-16 note representative= seed).by evaluating the score for the at least one concept of each document for a best to the clusters and assigning each document to the cluster with the best fit; and (See page 19 lines 4-10). While Lind does not differ substantially from the claimed invention the

disclosure of a threshold module determining the similarity between each of the documents grouped into each cluster based on the center of the cluster and the scores assigned to each of the at least one concepts in that document dynamically determining a threshold for each cluster as a function of the similarity between each of the documents, and identifying and reassigning each of the documents having the similarity falling outside the threshold are not necessarily explicit. Dhill however does disclose a threshold module determining the similarity between each of the documents grouped into each cluster based on the center of the cluster and the scores assigned to each of the at least one concepts in that document dynamically determining a threshold for each cluster as a function of the similarity between each of the documents, (See column 3 lines 55-60 and column 5 line 55- column 6 line 5) and identifying and reassigning each of the documents having the similarity falling outside the threshold (See column 3 lines 60-65). It would have been obvious to an artisan of ordinary skill in the pertinent at the time the invention was made to have incorporated the teaching of Dhil into the system of Lind.The modification would have been obvious because the two references are concerned with the solution to problem of efficient document scoring and clustering, therefore there is an implicit motivation to combine these references. In other words, the ordinary skilled artisan, during his/her quest for a solution to the cited problem, would look to the cited references at the time the invention was made. Consequently, the ordinary skilled artisan, would have been motivated to combine the cited references since Dhil's teaching would enable Lind's users to reclassify documents based on the center of the cluster..

As for claim 2 the rejection of claim 1 is incorporated, and further Lind discloses: the scoring module calculating the score as a function of a summation of at least one of the frequency of occurrence, the concept weight, the structural weight, and the corpus weight of the at least one concept (See Page 23 lines 1-4).

As for claim 3 the rejection of claim 2 is incorporated, and further Lind discloses: a compression module compressing the score through logarithmic compression (See page 17 line 30-34).

As for claim 4 the rejection of claim 1 is incorporated, and further Lind discloses: the scoring module calculating the concept weight as a function of a number of terms comprising the at least one concept (See page 21 lines 25-28).

As for claim 5 the rejection of claim 1 is incorporated, and further Lind discloses: the scoring module calculating the structural weight as a function of a location of the at least one concept within the at least one such document (See page 18 lines 10-14).

As for claim 6 the rejection of claim 1 is incorporated, and further Lind discloses: the scoring module calculating the corpus weight as a function of a reference count of the at least one concept over the plurality of documents (See page 18 lines 19- 21 note: this is an inverse weight of the reference count).

Claims 9-14 are method claims corresponding to system claims 1-6 respectively, and are thus rejected for the reasons set forth in the rejection of claims 1-6.

Claim 17 is rejected for the same reasons as claim 9.

As for claim 18 Lind discloses: a scoring module scoring a document in an electronically-stored document set comprising: a frequency module determining a frequency of occurrence of at least one concept within a document (See page 18 lines 1-3); and a concept weight module analyzing a concept weight reflecting a specificity of meaning for the at least one concept within the document (See page 25 lines 27-30 note: rtc(t,c) is a value based on meaning); a structural weight module analyzing a structural weight reflecting a degree of significance based on structural location within the document for the at least one concept (See page 18 lines 8-13), a corpus weight module analyzing a corpus weight inversely weighing a reference count of occurrences for the at least one concept within the document (See page 18 lines 19- 21 note: this is an inverse weight of the reference count); and a scoring evaluation module evaluating a score to be associated with the at least one concept as a function of the frequency, concept weight, structural weight, and corpus weight; (See page 21 24-27) and

A vector module forming the score assigned to the at least one concept as a normalized score vector for each such document in the electronically-stored document set, and a determination module determining a similarity between the normalized score

vector for each such document as an inner product of each normalized score vector (See page 30 line 30- page 31 line 1). A clustering module grouping the documents by the score into a plurality of clusters comprising; a selection submodule evaluating a set of candidate seed documents selected from the electronically-stored document set; a cluster seed submodule identifying seed documents by applying the similarity as a best fit to each such candidate seed document; an identification submodule identifying a plurality of non-seed documents; a comparison submodule determining the similarity between each non-seed document and a center of each cluster; and a clustering submodule assigning each non-seed document to the cluster with the best fit, subject to a minimum fit (See page 28 line 8-16 note representative= seed, (See column 5 lines 35-42).

While Lind does not differ substantially from the claimed invention the disclosure of a threshold module relocating outlier documents, comprising determining the similarity between each of the documents groups into each cluster based on the center of the cluster and the scores assigned to each of the at least one concepts in that document , dynamically determining a threshold for each cluster as a function of the similarity between each of the documents, and identifying and reassigning each of the documents with the similaritiy falling outside the threshold are not necessarily explicit. Dhill however does disclose a threshold module determining the similarity between eac of the documents grouped into each cluster based on the center of the cluster and the scores assigned to each of the at least one concepts in that document dynamically determining a threshold for each cluster as a function of the similarity between each of

the documents (See column 3 lines 55-60 and column 5 line 55- column 6 line 5); and identifying and reassigning each of the documents having the similarity falling outside the threshold (See column 3 lines 60-65).It would have been obvious to an artisan of ordinary skill in the pertinent at the time the invention was made to have incorporated the teaching of Dhil into the system of Lind.The modification would have been obvious because the two references are concerned with the solution to problem of efficient document scoring and clustering, therefore there is an implicit motivation to combine these references. In other words, the ordinary skilled artisan, during his/her quest for a solution to the cited problem, would look to the cited references at the time the invention was made. Consequently, the ordinary skilled artisan, would have been motivated to combine the cited references since Dhil's teaching would enable Lind's users to reclassify documents based on the center of the cluster..

As for claim 19 the rejection of claim 18 is incorporated and further Lind discloses: the scoring module evaluating the scoire in accordance with the formula Si = $\sum$ fij x cwij x swij x rwij where si comprises the score, fij comprises the frequency, 0<cwij $\leq$1 comprises the concept weight, o <swij $\leq$1 comprises the structural weight, and 0 < rwij $\leq$ 1 comprises the corpus weight for occurrence j of concept I (See page 23 lines 1-4).

As for claim 20, the rejection of claim 19 is incorporated and further Lindh discloses: the concept weight module evaluating the concept weight in accordance with the formula:

Cwij = 0.25 + (0.25 x tij), $1 \leq tij \leq 3$

0.25 + (0.25 x [7-tij]) $4 \leq tij \leq 6$

0.25, $tij \geq 7$ (See page 17 lines 30-34)

As for claim 21, the rejection of claim 19 is incorporated, and further Lindh discloses: the structural weight module evaluating the structural weight in accordance with the formula:

Swij = 1.0, if (J ≈ SUBJECT)

.8, if (J≈ HEADING)

.7, if (J≈ SUMMARY)

.5, if(J ≈ BODY)

.1, if (J≈ SIGNATURE)

where swij comprises the structural weight for occurrence j of each such concept I (See page 21 lines 25-29).

As for claim 22, the rejection of claim 19 is incorporated, and further Lindh discloses: the corpus weight module evaluating the corpus weight in accordance with the formula:

$$Rwij = \begin{cases} \left(\frac{T-rij}{T}\right)^2, & rij > M \\ 1.0, & rij \leq M \end{cases}$$

Where rwij comprises the corpus weight rij comprises a reference count for occurrence j of each such concept I, T comprises a total number of reference counts of documents in the document set, and M comprises a maximum reference count of documents in the document set (See page 23 lines 20-23).

As for claim 23, the rejection of claim 19 is incorporated and further Lindh discloses: a compression module compressing the score in accordance with the formula S`I = log(Si +1), where Si comprises the compressed score for each such concept I (See page 27 lines 1-7).

Claims 35-40 are method claims comprising substantially the same limitation as system claims 18-23, and are thus rejected for the reasons set forth in the rejection of claims 18-23.

Claim 52 is rejected for substantially the same reasons as claim 35,.

Claim 53 is an apparatus claim corresponding to method claim 18 and is thus rejected for the same reasons as claim 18.

Claims 24-27 and 41-44 and claims is rejected under 35 U.S.C. 103(a) as being unpatentable over Lind and Dhill as applied to claim 18 and 35 above, and further in view of US 6675159 (hereinafter Klein) (art of record)

As for claim 24 the rejection of claim 18 is incorporated, and further Klein discloses: a global stop concept vector cache maintaining concepts and terms (See column 18 lines 17-20 and See column 14 lines 45-49); and a filtering module filtering selection of the at least one concept based on the concepts and terms maintained in the global stop concept vector cache (See column 14 lines 45-50). It would have been obvious to an artisan of ordinary skill in the pertinent art at the time of the invention to have incorporated the teachings of Klein into the system of Lind. The modification would have been obvious because queries and documents are linked in the fact that words are the entities that are being processed. Therefore, any transformation capable of being made to a query should be able to applied to documents too, this makes all document management systems more efficient and easier to maintain.

As for claim 25 the rejection of claim 18 is incorporated, and further Klein discloses: a parsing module identifying terms within at least one document in the

document set, and combining the identified terms into one or more of the concepts (See column 2 lines 53-56).

As for claim 26 the rejection of claim 25 is incorporated, and further Klein discloses: the parsing module structuring each such identified term in the one or more concepts into canonical concepts comprising at least one of word root, character case, and word ordering (See column 14 lines 63-67).

As for claim 27 the rejection of claim 25 is incorporated, and further Klein discloses wherein at least one of nouns, proper nouns and adjectives are included as

Claims 41-44, are method claims corresponding to system claims 24-27, respectively and are thus rejected for the same reasons as set forth in the rejection of claims 24-27,.

Claims 30,31,47 and 48 are rejected under 35 U.S.C. 103(a) as being unpatentable over Lind and Dhill as applied to claim 29 above, and further in view of Klein.

As for claim 30 the rejection of claim 29 is incorporated, and further Klein discloses: a normalized score vector for each document comprising the score

associated with the at least one concept for each such concept occurring within the document (See column 3 lines 18-21); and the similarity module determining the similarity as a function of the normalized score vector associated with the at least one concept for each such document (See column 18 lines 23-26).

As for claim 31, the rejection of claim 30 is incorporated, and further Klein discloses: the similarity submodule calculating the similarity in accordance with the formula

$$\cos\theta\ ab = \frac{(Ss \bullet Sb)}{Sa\ Sb}$$

Where cosθ ab comprises a similarity between a document A and a document B, Sa comprises a score vector for document A and Sa comprises a score vector for document B.

Claim 48 is a method claim corresponding to the system of claim 31 respectively and is thus rejected for the same reasons as set forth in the rejection of claim 31.

Response to Arguments

Applicant's arguments filed 9/7/2007 have been fully considered but they are not persuasive.

Applicant argues:

Further, the Lindh-Dhillon combination fails to teach assigning non-seed documents into a cluster with a best fit, subject to a minimum fit. Documents can be clustered using a clustering algorithm, such as k-means clustering (Lindh, p. 28, lines 9-11). A set of clusters containing similar documents will be produced (Lindh, p. 28, lines 6-7). Thus, each document will be clustered with similar documents based on a particular algorithm without applying further requirements, such as a minimum fit criterion. Therefore, the combination teaches assigning documents to clusters using a clustering algorithm, rather than grouping a non- seed document into a cluster with a best fit, subject to a minimum fit.

Examiner responds:

Examiner is not persuaded. Examiner is entitled to give claim limitations their broadest reasonable interpretation in light of the specification. Interpretation of Claims-Broadest Reasonable Interpretation. During patent examination, the pending claims must be 'given the broadest reasonable interpretation consistent with the specification.' Applicant always has the opportunity to amend the claims during prosecution and broad interpretation by the examiner reduces the possibility that the claim, once issued, will be

interpreted more broadly than is justified. In re Prater, 162 USPQ 541,550-51 (CCPA 1969). Accordingly the documents in Lind-dhillon are best fit.

Conclusion

THIS ACTION IS MADE FINAL. Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the mailing date of this final action.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Leon J. Harper whose telephone number is 571-272-0759. The examiner can normally be reached on 7:30AM - 4:00Pm.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Hosain T. Alam can be reached on 571-272-3978. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

LJH
Leon J. Harper
November 22, 2007

HOSAIN ALAM
SUPERVISORY PATENT EXAMINER